Question-Answer system using NLP: An application of BERT

What is NLP ?

- Computers speak in binary (0 and 1) because of how they are built.
- In simple words, teaching human language to computers
- Standard definition -

Natural language processing NLP is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data.(wikipedia)

Why need NLP ?

- To make computers/machines more accessible
- NLP has many real life applications like -
    - Voice assistants(siri,alexa,cortana etc)
    - Spam filtering,smart reply
    - Recommendation systems(used on amazon, IMdb, for books etc)
    - Sentiment analysis(customer reviews on products)
    - Language translators

NLP Pipeline

NLP pipeline - A set of steps to build an end to end nlp software

Steps in NLP pipeline -

- Data Acquisition - We need data to build nlp applications
- Text preparation
- Feature engineering
- Modeling - Applying different types algorithms on our data depending on our purpose
- Deployment - Deploying our application on a website or cloud service etc.

Text preparation

Following processes are present in text preparation -

- lowercasing the text
- remove punctuations
- removing stop words
- tokenization
- stemming
- POS tagging
- remove emojis
- spell check

All these procedures can be carried out by python libraries like nltk

Feature engineering

We can't apply any algorithm on textual data

We have to convert words/sentences/paragraphs in input data to numbers in a meaningful way.(we can’t just assign random numbers)

There are several methods to convert text into numbers -

1. Bag of words
2. TF-IDF
3. Word2Vec

01 Bag of words

Converting sentences into vectors

Let's say we have 3 sentences -

1. He is a good boy
2. She is a good girl
3. Boy and girl are good

After text preparation we will get something like -

1. good boy
2. good girl
3. boy girl good

Now let's create a matrix-like structure where rows are sentences in a data and columns are all the different words present in a data. -

	boy	girl	good
Sentence 1(good boy)	1	0	1
Sentence 2(good girl)	0	1	1
Sentence 3(boy girl good)	1	1	1

Now we got a vector for each sentences -

1. He is a good boy - [1,0,1]
2. She is a good girl - [0,1,1]
3. Boy and girl are good - [1,1,1]

In bag of words no semantic meaning is captured

02 TF-IDF

TF-IDF stands for term frequency-inverse document frequency

It converts words into vectors

Say we have dataset D with 3 sentences d1,d2,d3 -

TF(t,d) = (# occurrences of t in d)/total no. of words in d (t - word , d - sentence)

IDF(t,D) = log((total sentences in D)/(# sentences with t))

TF-IDF = TF(t,d)xIDF(t,D)

03 Word2Vec

Developed by Google engineers

Word2Vec is pretrained model It is trained on google News dataset

Word2Vec converts words into vectors

Word2vec captures the semantic meaning of words

Word2vec converts words into low dimensional vectors as compared to TF-IDF and bag of words

Word2Vec gives Dense vectors(most of entries are non zero)

How Word2Vec captures semantic meaning ?

The underlying assumption of Word2Vec is that 2 words sharing similar contexts also share a similar meaning and consequently similar vector representation.

Word2Vec will try to create features to identify each word.

Features - Like for word “human” we can say features are gender,weight,height etc

	King	Queen	Man	Woman	Monkey
Gender	1	0	1	0	1
Wealth	1	1	0.3	0.3	0
Power	1	0.7	0.2	0.2	0
Weight	0.8	0.4	0.6	0.5	0.3
Speak	1	1	1	1	0

Now if we do -

King - Man + Woman

= [1,1,1,0.8,1] - [1,0.3,0.2,0.6,1] + [0,0.3,0.2,0.5,1]

= [0,1,1,0.7,1]

≈ Queen

As we can see the resultant vector and vector for queen are similar.

Now as we can see the machine is able to see the meaning of words and is able to do the above word math.

Here for words we put the weight values by our knowledge but this weight will be decided by Word2Vec model

How Word2Vec finds the weights for the words ?

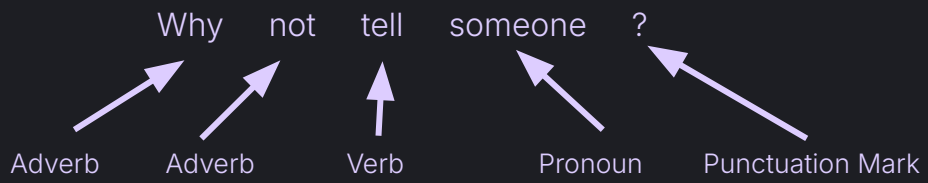
We take a fake problem and in the process of solving this problem we get a vector for a word as a byproduct.

There are 2 methods of doing this -

1. CBOW - continuous bag of words
2. Skip-gram

Part of Speech Tagging

A task of labelling each word in a sentence with its appropriate part-of-speech.



Part of Speech Tagging is a data pre-processing step.

Applications

Named Entity Recognition

It is an information retrieval process.

Question-Answering System

POS tagging is used as a pre-processing step while building question-answering systems.

Word Sense Disambiguation

If one word is being used in two different sentences with different meaning, one can tell difference between them after POS tagging.

Chatbots

Working Principle

Computer uses Hidden Markov Model (HMM) algorithm for doing POS tagging.

To reduce the computational complexity of POS Tagging, *Viterbi Algorithm* is used.

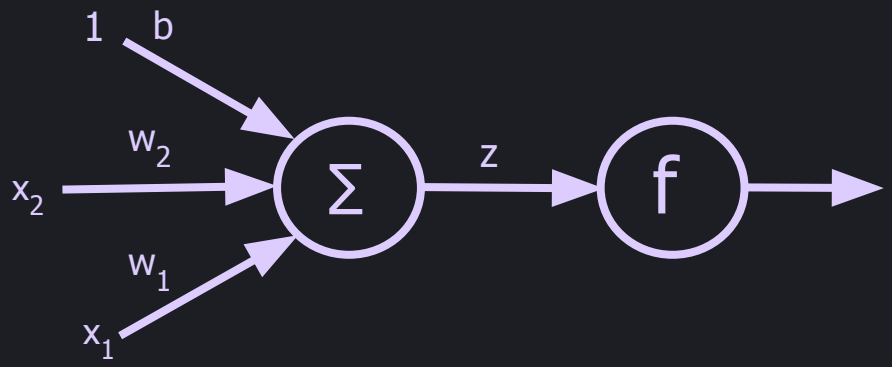
Perceptron

What is perceptron ?

It is an algorithm

It is used for supervised machine learning

We can see perceptron as a mathematical model/function



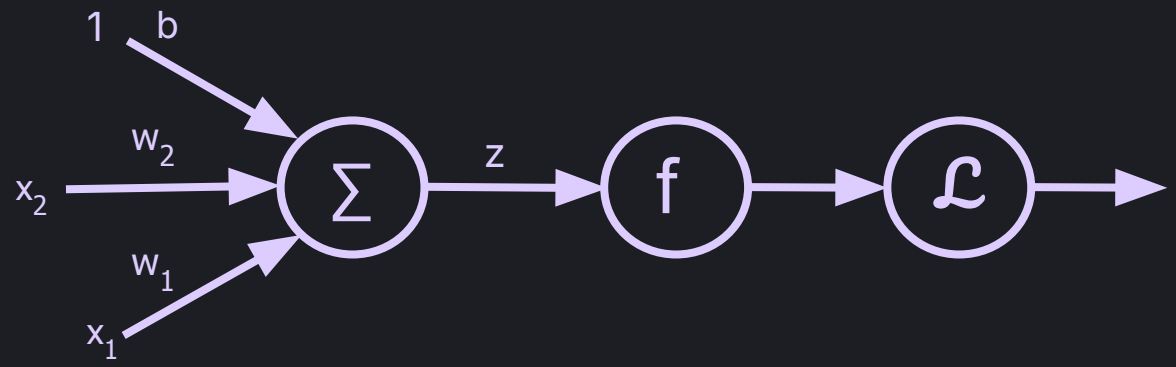
$Z = w_1x_1 + w_2x_2 + b$

Activation function converts output(z) in a particular range

In simple words activation function interprets the result(z)

Some examples of activation function are -

1. Step-function
2. ReLU
3. Tanh
4. Sigmoid



Loss Function	Activation function	Output
Hinge Loss	Step function	perceptron - binary classifier
log-loss(binary cross entropy)	sigmoid	Logistic regression - binary classifier
Categorical cross entropy	softmax	softmax regression - multiclass classifier
MSE(mean square error)	Linear(No activation function)	Linear regression

Problem with perceptron

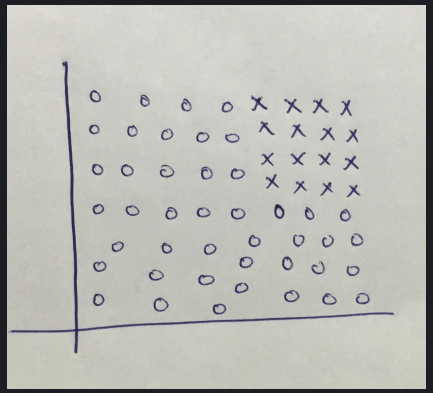
Perceptron can’t be fitted on a data which is not linearly separable.

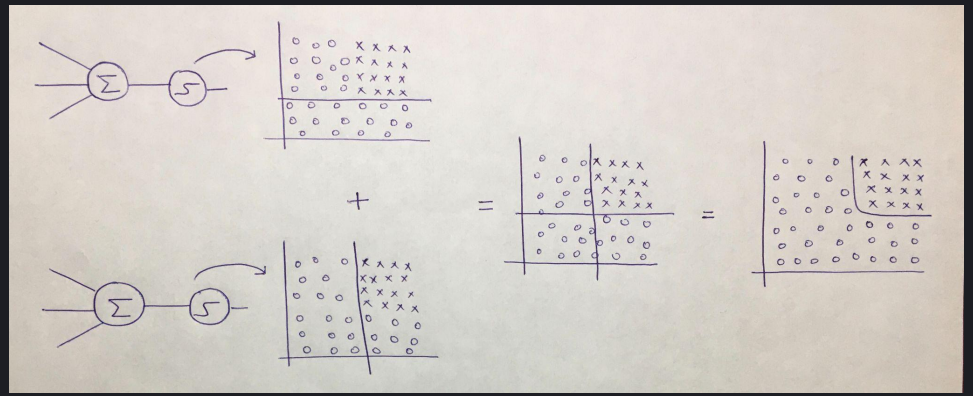
Which means if one line or plane or hyperplane can’t separate your data then we can’t fit perceptron on it

Click Here

Multilayer Perceptron(ANN)

How does multilayer perceptron captures non-linearity ?





Click Here

RNN-Recurrent Neural Network

What is RNN ?

It is special type of neural network

It is sequential model and is used to work on sequential data

Sequential data - data where sequence matters

1. Text
2. Time series data

Why not ANN ?

In ANN we feed all the information at once therefore important sequential information is lost

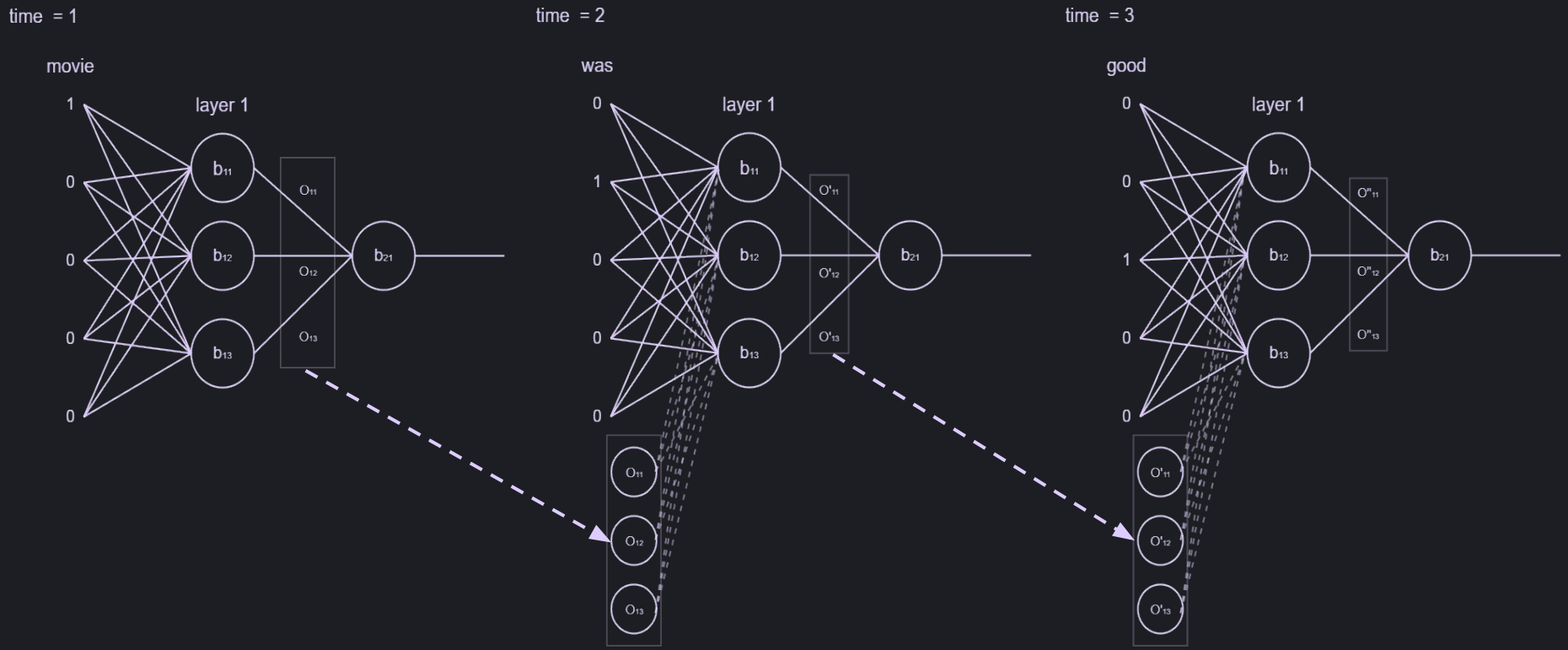
ANN do not perform well on sequential data

Textual data can be of different sizes and we can’t vary the input size in ANN

RNN is being heavily used in natural language processing

Applications of RNN -

- Sentiment analysis
- Sentence completion
- Image caption generation
- Google translate


time = 1
movie
layer 1
1
0
0
0
0
b11
b12
b13
O11
O12
O13
b21
time = 2
was
layer 1
0
1
0
0
0
b11
b12
b13
O'11
O'12
O'13
b21
O11
O12
O13
time = 3
good
layer 1
0
0
1
0
0
b11
b12
b13
O"11
O"12
O"13
b21
O'11
O'12
O'13

Transfer Learning

Deep learning models are data hungry

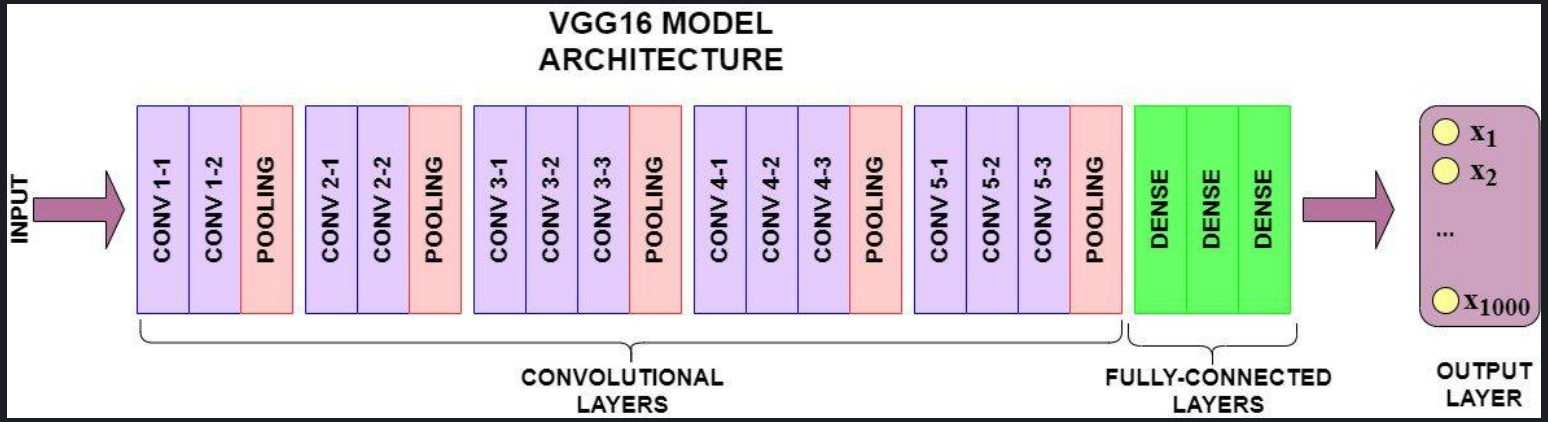
Model can take lot of time to train

Using pre-trained models is solution to this problem

But pre-trained model will not be very useful if our task is very specific

Standard definition -

Transfer learning (TL) is a research problem in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.(wikipedia)



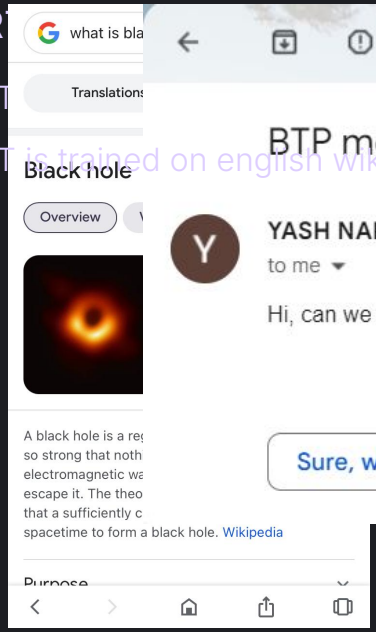
VGG16 model is trained on ImageNet dataset which has around 1000 different classes.

BERT

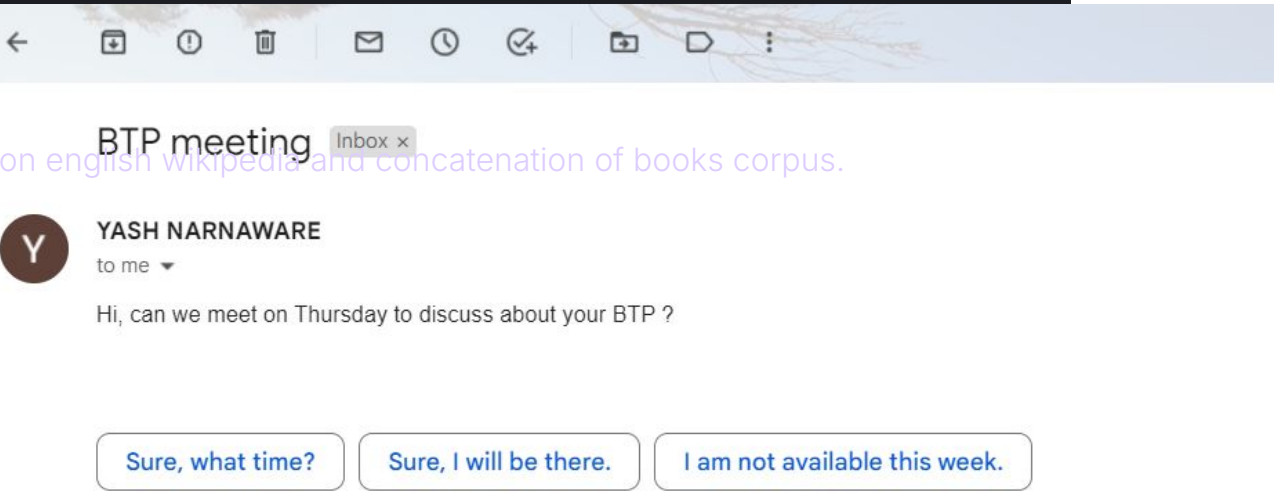
BERT

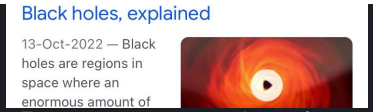
BERT

BERT is trained on english wikipedia and concatenation of books corpus.



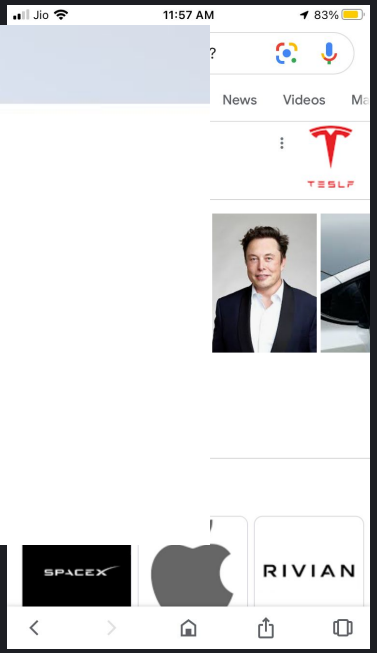
Text Summarization



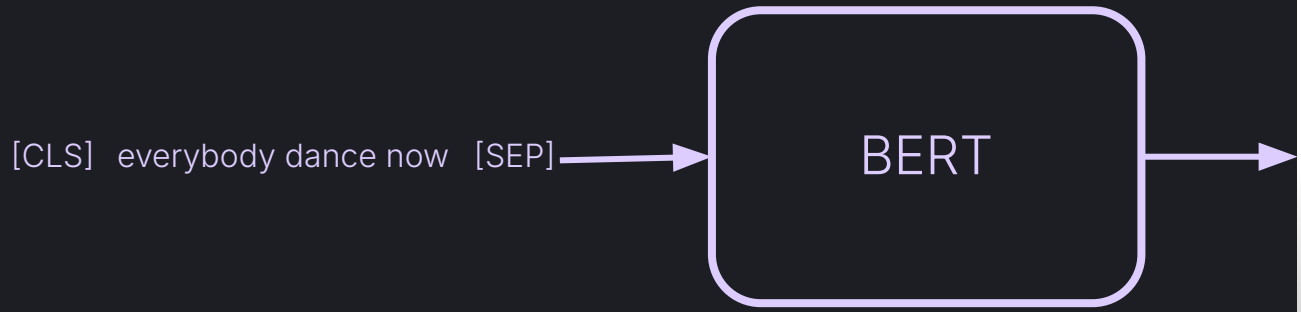


Response selection

Text encoding
similarity retrieval

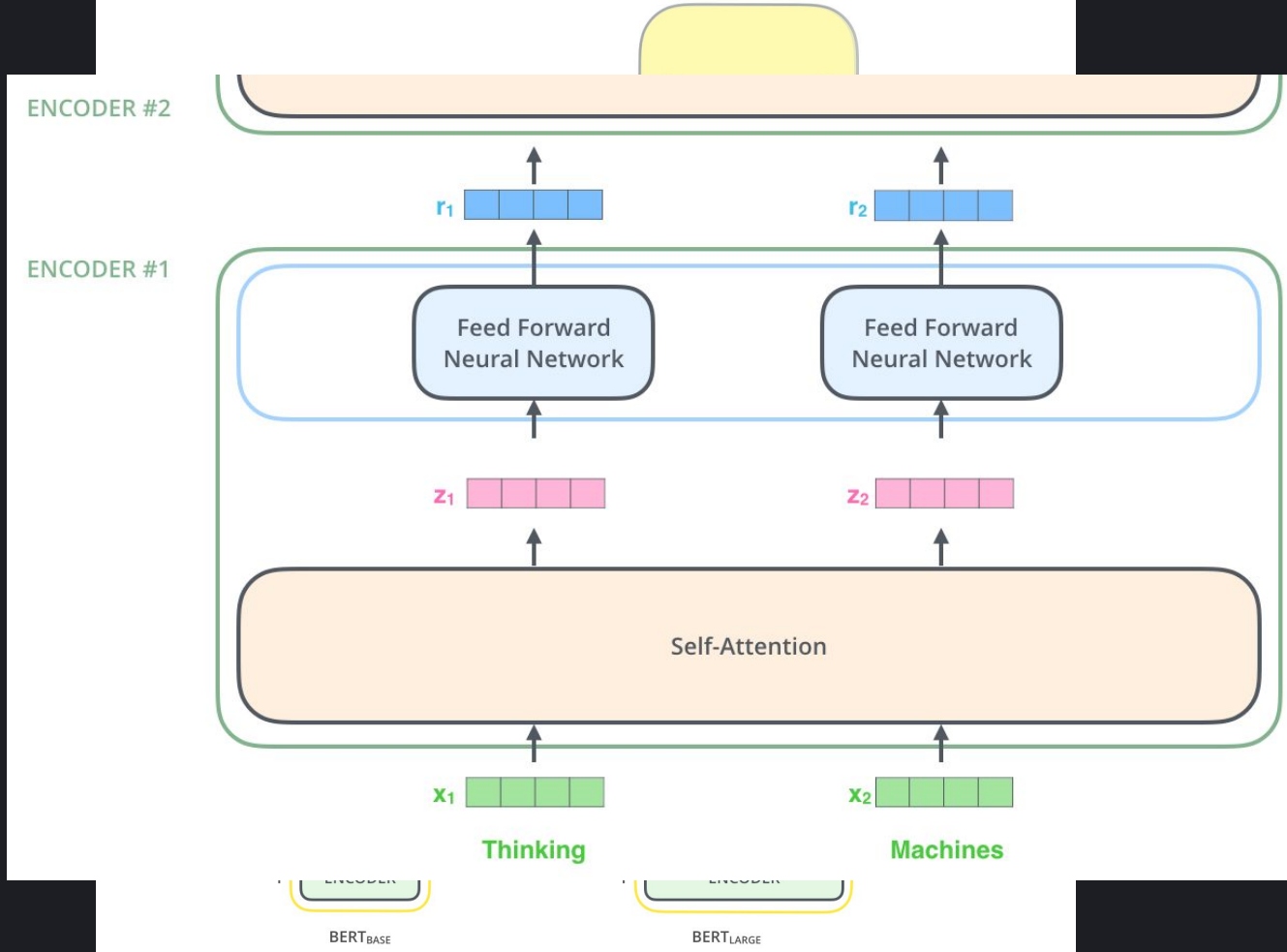


Question Answering

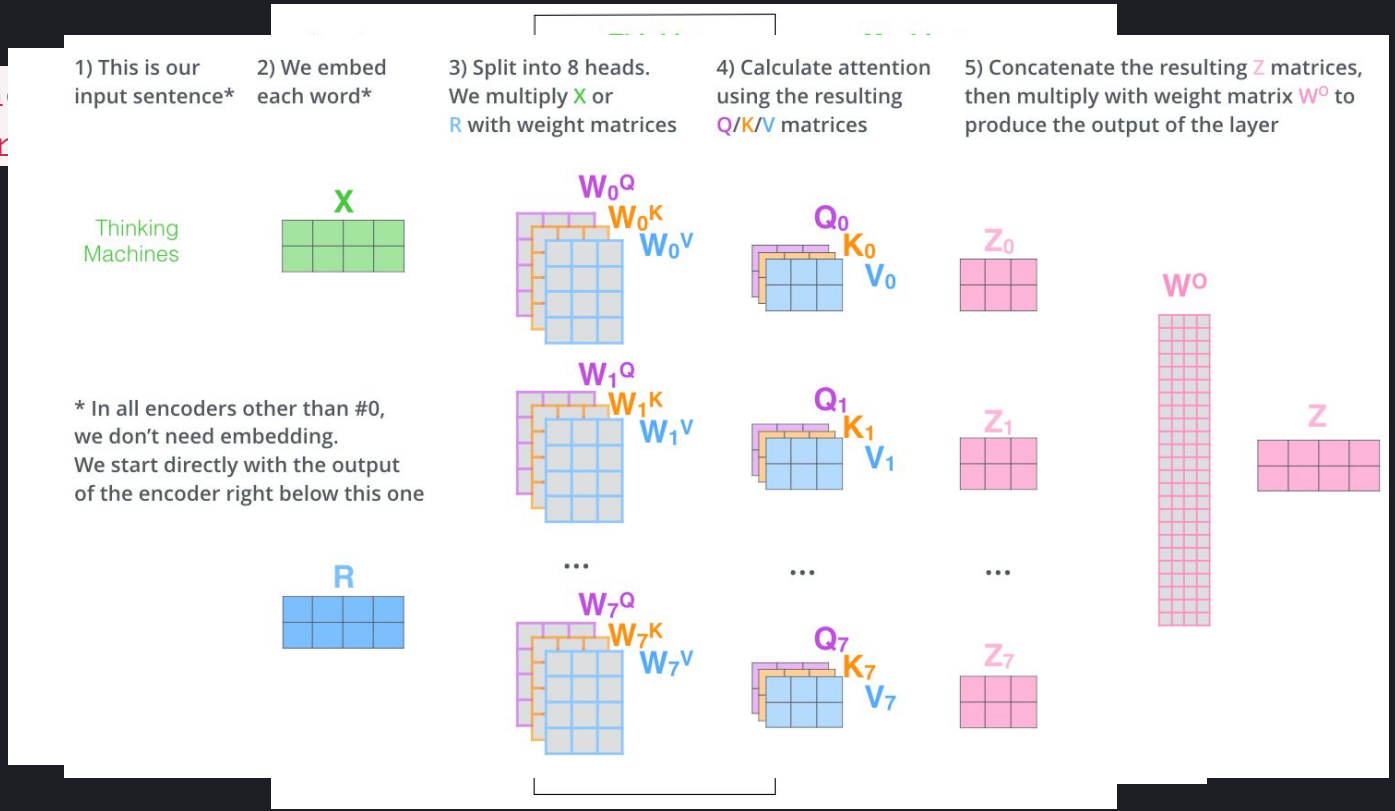


	[CLS]	everybody	dance	now	[SEP]
0	0.115728	-0.150873	0.076831	-0.661984	0.834172
1	0.223339	0.269943	0.009255	-0.858786	0.279408
2	0.394565	0.411511	1.518060	0.053227	0.127079
3	0.021645	-0.369215	-0.217087	-0.251777	0.270489
4	-0.569364	-0.537542	-0.033147	0.340954	-0.452304
...	...	...	...	...	...
763	0.454554	0.539476	0.672731	0.648410	-0.864230
764	-0.177935	-0.192709	-0.515615	-0.102858	0.159125
765	-0.570071	0.058834	-0.369875	0.195014	-0.065941
766	-0.257792	0.029307	-0.338636	-0.121347	-0.979295
767	0.380097	0.632699	0.124689	-0.666084	-0.147284

768 rows × 5 columns


ENCODER #2
ENCODER #1
r1
r2
Feed Forward Neural Network
Feed Forward Neural Network
z1
z2
Self-Attention
x1
x2
Thinking
Machines
BERTBASE
BERTLARGE

What is self attention ?



Word2Vec based Bhagavad Gita verse recommender

The dataset looks like this -

	verse_number	verse
0	Bhagavad Gita 1.1	Dhritarashtra said: O Sanjay, after gathering ...
1	Bhagavad Gita 1.2	Sanjay said: On observing the Pandava army sta...
2	Bhagavad Gita 1.3	Duryodhan said: Respected teacher! Behold the ...
3	Bhagavad Gita 1.4 – 1.6	Behold in their ranks are many powerful warrio...
4	Bhagavad Gita 1.7	O best of Brahmins, hear too about the princip...
...	...	...
635	Bhagavad Gita 18.74	Sanjay said: Thus, have I heard this wonderful...
636	Bhagavad Gita 18.75	By the grace of Veda Vyas, I have heard this s...
637	Bhagavad Gita 18.76	As I repeatedly recall this astonishing and wo...
638	Bhagavad Gita 18.77	And remembering that most astonishing and wond...
639	Bhagavad Gita 18.78	Wherever there is Shree Krishna, the Lord of a...

640 rows × 2 columns

Word2Vec based Bhagavad Gita verse recommender

```
import gens
from gensim                                                    Vectors

model = KeyedVectors.load_word2vec_for                         00.bin.gz',binary=True,limit=500000)
```

```
def preprocess_and_vectorize(text):

    text = text.lower()
    text = nltk.word_tokenize(text)

    temp = []
    for i in text:
        if i.isalnum() == True :
            temp.append(i)

    temp1 = []
    for i in temp:
        if i not in stop_words:
            temp1.append(i)

    temp2 = []
    for i in temp1:
        if i in model:
            temp2.append(model[i])

    temp3 = np.mean(temp2, axis=0)
    return(temp3)
```

Word2Vec based Bhagavad Gita verse recommender

```
user_input = input("Enter the text = ")
user_input = preprocess_and_vectorize(user_input)
```

```
Enter the text = how to achieve success ?
```

```
Bhagavad Gita 6.45 With the accumulated merits of many past births, when these yogis engage in sincere endeavors to make furt
her progress, they become purified from material desires and attain perfection in this life itself.


Bhagavad Gita 6.36 Yog is difficult to attain for one whose mind is unbridled. However, those who have learnt to control the
mind, and who strive earnestly by proper means, can attain perfection in Yog. This is My opinion.


Bhagavad Gita 5.5 The supreme state that is attained by means of karm sanyās is also attained by working in devotion. Hence,
those who see karm sanyās and karm yog to be identical, truly see things as they are.


Bhagavad Gita 8.15 Having attained Me, the great souls are no more subject to rebirth in this world, which is transient and f
ull of misery, because they have attained the highest perfection.


Bhagavad Gita 5.4 Only the ignorant speak of sānkhya (renunciation of actions, or karm sanyās) and karm yog (work in devotio
n) as different. Those who are truly learned say that by applying ourselves to any one of these paths, we can achieve the res
ults of both.
```

BERT based QA system

```
from transf
tokenizer =
```

```
ned-squad')
```

```
from transfo
model = Bert
```

```
uned-squad')
```

IIT Goa is offering Under Graduate Programmes from the year 2016. The broad objectives of the programme are:

- To provide the highest level of education in the field of technology and science, and to produce competent engineers and scientists with well-honed managerial and entrepreneurial skills, having team spirit and leadership qualities who can contribute towards the Country and the Society
- To promote a spirit of free and objective inquiry, and development of knowledge in different disciplines.
- To increase student participation in Nation building through technology development that is sensitive to local needs.

Academic Programmes are governed by Rules and Regulations approved by the Senate from time to time. The Senate is a statutory and supreme body of the Institute that governs all academic matters of the Institute, and the rulings of Senate Chairman are final in regard to all academic issues. The Senate continuously monitors the academic Programmes and makes appropriate modifications /improvements as and when required. The Senate also sets a definite time schedule for various academic activities. This manual sets out the procedures and requirements of the undergraduate Programmes of study that fall under the purview of the Senate Under-Graduate Committee (SUGC) a standing committee of the Senate.

Who makes the rules of UG program ?

```
question = input("Enter the question - ")
```

Answer: "the senate"

From when IIT goa is offering UG program ?

```
answer_question(question, data)
```

What is full form of SUGC ?

Answer: "2016"

Answer: "senate under - graduate committee"

Who makes the final call in academic issues ?

Answer: "senate chairman"

Thank You